Available online at www.sciencedirect.com

ScienceDirect

Genomics 91 (2008) 203–208

GENOMICS

www.elsevier.com/locate/ygeno

Analysis of *PKD1* for genomic deletion by multiplex ligation-dependent probe assay: Absence of hot spots

Piotr Kozlowski [a,*], John Bissler [b], York Pei [c], David J. Kwiatkowski [a]

[a] *Translational Medicine Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA*
[b] *Division of Nephrology and Hypertension, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA*
[c] *Division of Nephrology and Division of Genomic Medicine, Department of Medicine, University Health Network and University of Toronto, Toronto, ON, Canada*

Received 18 July 2007; accepted 3 October 2007
Available online 3 December 2007

Abstract

Autosomal dominant polycystic kidney disease is largely due to mutations in *PKD1*. *PKD1* has an unusual genomic structure, including a 2.5-kb polypyrimidine sequence in intron 21, which has been postulated to lead to a high rate of spontaneous genomic mutation events. In addition, the majority of the gene is duplicated three to six times at 97–99% identity elsewhere in the genome. To identify genomic mutations in *PKD1,* we developed a multiplex ligation-dependent probe assay (MLPA) in which sites of variation between *PKD1* and its copies were positioned at the ligation sites of the MLPA probe sets. Thirteen probe sets covered *PKD1* exons 2 through 46, at an average spacing of 2.5 kb. Analysis of 27 independent PKD patient samples showed no evidence for genomic deletions confined to *PKD1.* Analysis of 15 tuberous sclerosis patient samples in which deletions in *TSC2* extended into *PKD1* showed no evidence of clustering of breakpoints near the polypyrimidine tract.

Keywords: PKD1; TSC2; MLPA; Large deletions; Polypurine–polypyrimidine tract; Intron 21; Autosomal dominant polycystic kidney disease; Multiplex ligation-dependent probe amplification

Autosomal dominant polycystic kidney disease (ADPKD) is a common genetic disease and accounts for approximately 10% of end-stage renal disease in the United States [1]. ADPKD is caused by mutations in either the *PKD1* gene, which account for approximately 85% of ADPKD families (MIM 601313), or the *PKD2* gene, which account for about 15% of families (MIM 173910), based upon linkage analyses [2,3].

The *PKD1* gene has an unusual genomic structure (Figs. 1A and 2A). It consists of 46 exons, of which the last 45 are found in a relatively compact 31 kb on 16p13.3, directly adjacent to the *TSC2* gene. *PKD1* and *TSC2* are in opposite transcriptional orientations, such that their 3′ UT regions overlap by a few nucleotides [4]. The first exon of *PKD1* is located 16 kb away from exon 2, with a relatively large intron 1. Moreover, the region from intron 1 to exon 33 of *PKD1* is segmentally duplicated six times elsewhere on chromosome 16 at 97–99% identity [5,6]. In addition, *PKD1* exon 1 and the adjacent intron 1 sequence are segmentally duplicated three times at 99% identity.

These genomic features have greatly complicated mutation analysis of *PKD1* in ADPKD patients, necessitating complex design strategies, including long-range PCR preamplification using unique sequence elements and other approaches. With these measures, comprehensive mutation analyses of *PKD1* indicate that about 50–60% of patients meeting the diagnostic criteria for ADPKD will have a small (including indels of size <30 bp) mutation found in *PKD1* [7]. One recent small series was able to identify 100% of mutations in Finnish PKD patients, and 16 of 17 families showed both linkage to and mutation in *PKD1* [8].

Another highly unusual feature of the *PKD1* gene is the presence of a 2.5-kb polypurine–polypyrimidine sequence in intron 21 and another shorter, similar sequence in intron 20. The intron 21 sequence consists of 65% cytosine and 32% thymidine (97% pyrimidine) on the coding strand and is one of the largest intragenic tracts of this kind in the human genome. This sequence has been shown to form various types of non-B DNA including triplex DNA structures [9], confer plasmid instability in bacterial systems [10], and interfere with DNA replication [11].

* Corresponding author. Current address: Laboratory of Cancer Genetics, Institute of Bioorganic Chemistry, PAS, Noskowskiego 12/14, 61-704 Poznan, Poland. Fax: +48 061 8528919.
E-mail address: kozlowp@yahoo.com (P. Kozlowski).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.10.003

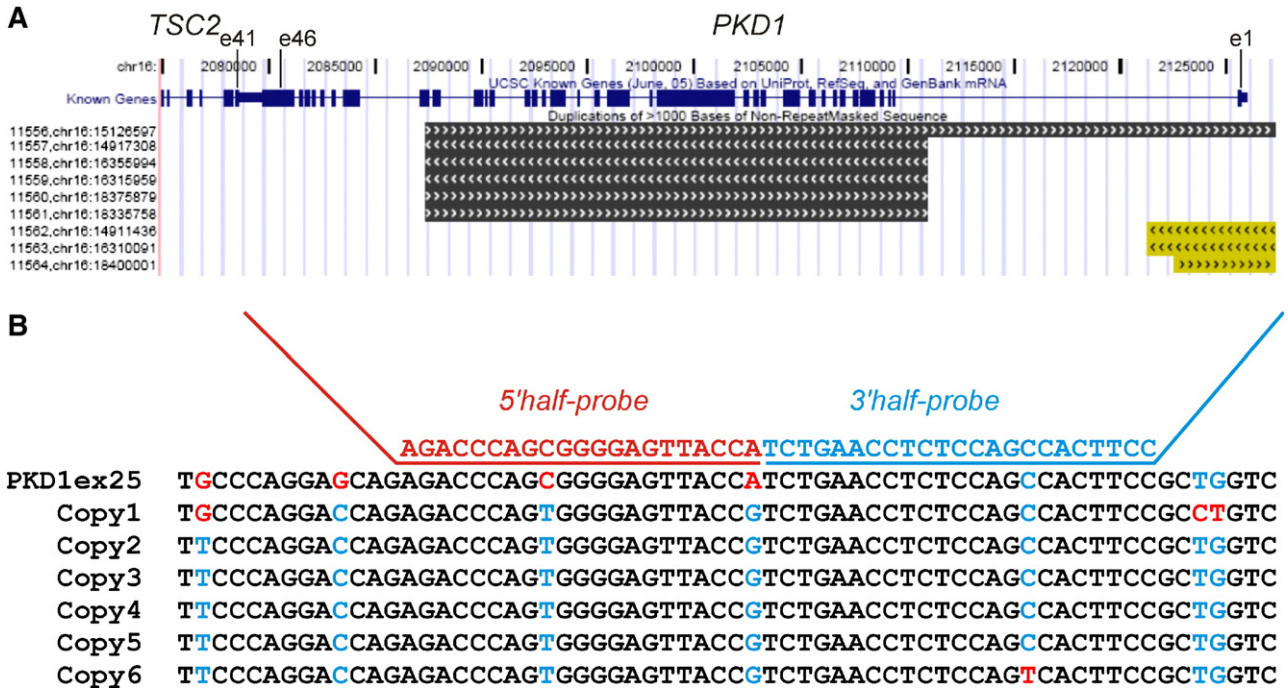

Fig. 1. Genomic structure of *PKD1* and strategy for probe design in duplicated regions. (A) Genomic structure of *PKD1,* according to the UCSC genome browser (Mar 2006), with first (e1) and last (e46) exons indicated. Black and yellow bars indicate *PKD1* fragments duplicated elsewhere in the genome at 97–98% and 98–99% identity, respectively. (B) Sequence alignment for *PKD1* exon 25 and six duplicated copies of that exon located elsewhere on chromosome 16. Blue and red highlight matching and discordant nucleotides. Top: 5′ and 3′ half-probes are shown. Note that one of the nucleotides at the ligation position (position −1 in the 5′ half-probe, A) is present only in the *PKD1* copy, not in the duplicates, by our design (additionally, the nucleotide at position −9 is *PKD1* specific).

These observations have led to speculation that this sequence would be associated with a higher than average rate of mutation events, particularly genomic deletions. Rossetti et al. noted an increase in point mutations in the vicinity of the tracts [7], and there appears to be increased recombination and mutagenic activity in the *PKD1* region on 16p13.1 [12–14]. However, limited surveys using Southern blot analysis have identified genomic deletions of *PKD1* in only a small fraction of ADPKD patients [5,15].

Here we present the development of a comprehensive multiplex ligation-dependent probe assay (MLPA) to assess deletions and duplications within the *PKD1* gene. Due to the extensive, nearly identical sequence repeats within *PKD1,* assay design was challenging. We then studied cohorts of ADPKD patients and a group of patients with the combined tuberous sclerosis (TSC)–PKD syndrome [16,17]. Fifteen patients with genomic deletions encompassing part of the *PKD1* gene were studied, and there was no evidence from the positions of the deletion breakpoints that the intron 21 polypurine–polypyrimidine tract played a role in the occurrence of these germ-line deletions.

Results and discussion

To examine the frequency of genomic deletion mutations in *PKD1* and localize the breakpoints within the gene for such deletions, we developed an MLPA to determine the copy number of multiple exons within *PKD1.* Thirteen MLPA probes were designed to cover exons 2 through 46 (3′ end) of the *PKD1* gene, with an average interprobe spacing of 2.5 kb (Fig. 2A). Several additional probes were also designed: 1 in *PKD1* intron 1, 2 in the 5′ upstream region of *PKD1,* and 2 in exons 4 and 6 of *PKD2.* We also included as controls three probe sets from other chromosomes [18]. The complete set of MLPA probes consisted of 22 probes whose amplification products ranged in size from 90 to 154 bp. The sequence of each probe and its exact position are shown in Supplementary Table 1 and Fig. 2A. To maximize the ability to discriminate between *PKD1* exons and duplicated segments, nucleotides that varied between these two sequences were placed in the probes at the ligation site (see Materials and methods for further details) (Fig. 1B).

Extensive analysis on normal samples was performed to demonstrate consistent signal intensity for these probes (Fig. 3A). There was no evidence of large unexpected peaks reflecting hybridization and subsequent amplification of probes at more than one site per genome. In the course of this analysis, 12 probe sets had to be modified or removed, due to variation in the size of amplification peaks among different DNA samples. Such variation in probe-set amplicon yield may have occurred for any of several reasons: (1) unknown sequence variation in the *PKD1* sequence chosen for primer design, (2) inaccuracies and/or sequence variation in the multiple copies of duplicated regions, or (3) variable numbers of repeats of some of these sequences.

We then analyzed 60 DNA samples from ADPKD patients from 27 families. In this set there were 4 samples that showed evidence of copy number variation (Fig. 3B). Three of these samples were from a single family, and all showed an apparent deletion of exon 40 (Fig. 3B, sample 01-1090). DNA sequence

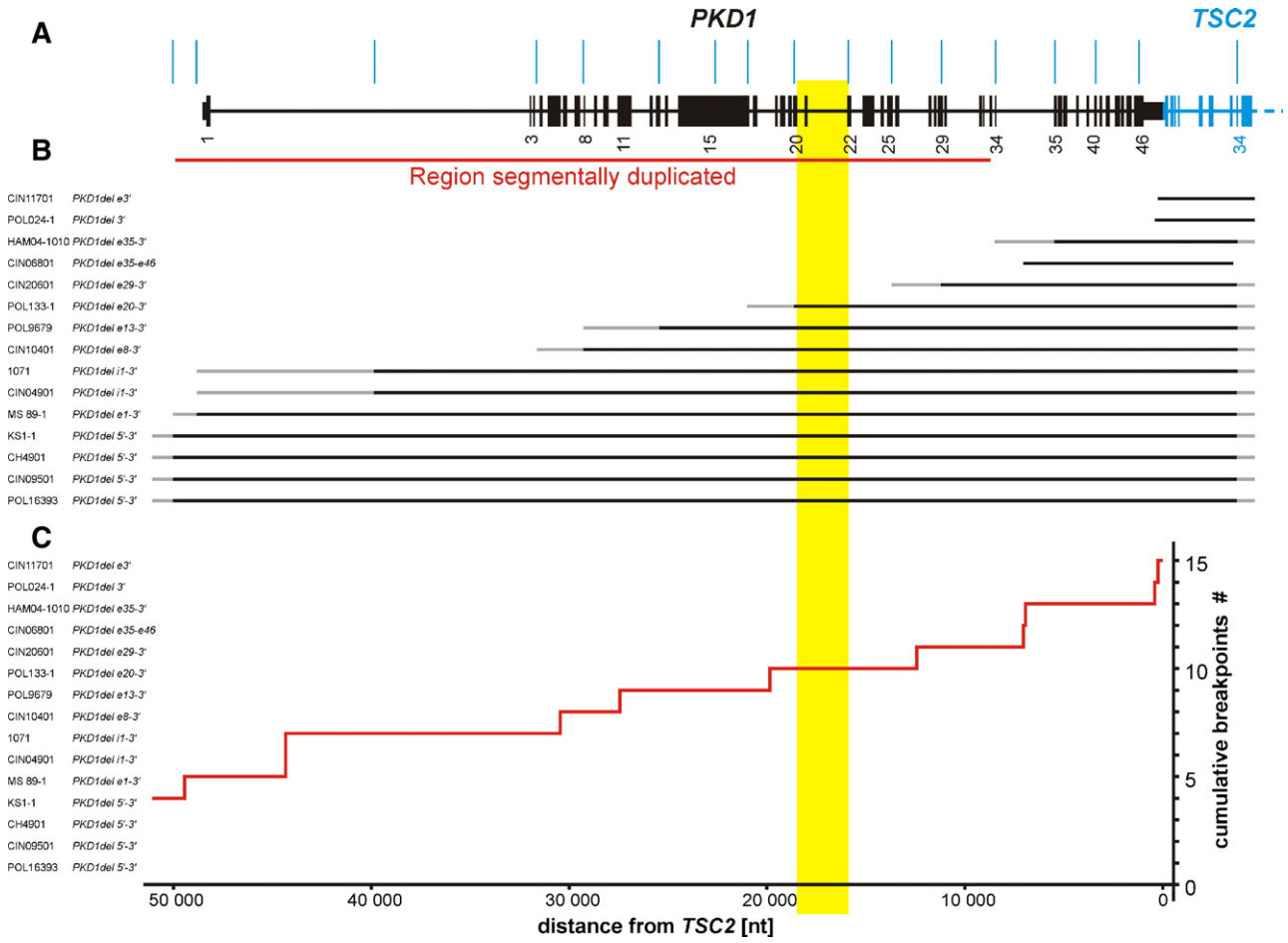

Fig. 2. Map of all *PKD1* deletions. (A) Map of *PKD1* and flanking regions. Exons are indicated by black vertical lines of width proportional to exon size. Thin blue lines indicate the positions of *PKD1* MLPA probes. The duplicated regions are indicated by the red line. (B) Deletion map. Each deletion mutation in *PKD1* is represented by a horizontal black/gray line indicating its extent, as mapped previously in *TSC2* [18] and in *PKD1* (this article). The black line indicates the minimum deleted region, and the gray the possible extent to the next undeleted probe. For two mutations the breakpoints were identified by LRPCR/sequencing. (C) Cumulative frequency of breakpoints within *PKD1* as a function of linear distance. The positions of mutations were assigned to the midpoint of the distance between the last probe showing deletion and the first probe not showing deletion. Note that breakpoints are distributed fairly evenly across *PKD1* with none in intron 21 or 22, highlighted with yellow background. Intron 21 contains the polypurine–polypyrimidine tract.

analysis of a PCR product from this exon demonstrated the presence of a previously unknown single-base deletion mutation, c.11372delC, which was situated under the oligonucleotide probe (9 bases from the ligation position), accounting for the reduction in signal intensity by MLPA. The other sample showed evidence of deletion for probes from *PKD1* intron 1 through exon 46 (Fig. 3B, sample 18-1071). This patient had features of TSC in addition to early onset ADPKD, consistent with previous observations on the effects of genomic disruption of both genes [17,18]. Thus, there were no intragenic genomic deletions of *PKD1* identified in this analysis.

We also examined 15 DNA samples from TSC patients who had previously been studied using MLPA probe sets for *TSC2* and shown to have deletions that extended into *PKD1* [18]. The extent of the deletions in *PKD1* were entirely consistent with our previous analysis, providing additional strong support for the validity of this new *PKD1* MLPA. These deletions were found to extend a variable extent into the *PKD1* gene (Figs. 2B, C, and 3B). Four deletions extended through the entire *PKD1* gene, including both probe sets in the 5′ flanking region of *PKD1*. The other 11 appeared to be randomly distributed across the length of the gene. In particular, none of the breakpoints were found in the region between the probes for exons 20 and 22, which contains the intron 21 polypurine–polypyrimidine tract.

In conclusion, we developed a reliable and robust MLPA analytic method for a region of the genome that is highly repetitive. Although other strategies may be considered, we found that designing the oligonucleotide probes to make the ligated base overlie the site of sequence differences among highly similar sequences was effective to permit discrimination of unique sequence copy number. We observed greater variation in the prenormalized peak heights (Fig. 3A) than seen in MLPA probe sets designed for other genes [18], possibly reflecting hybridization of probes to multiple sites in the genome or re-hybridization of *PKD1* exons to duplicated copies after denaturation. Nonetheless, the peak height patterns were highly consistent, enabling detection of deletion mutations.

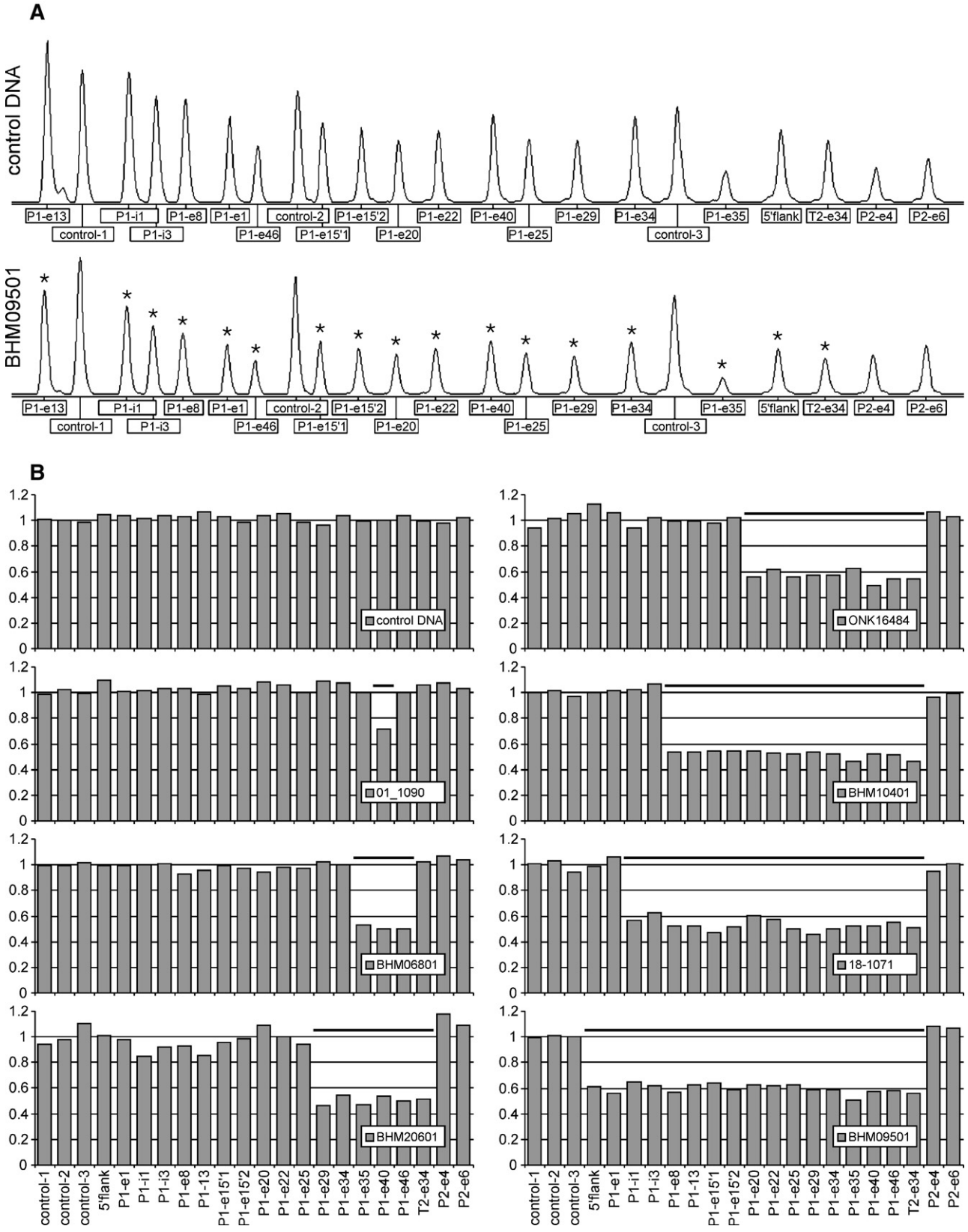

Fig. 3. Electropherograms and normalized peak height graphs for *PKD1* MLPA. (A) Electropherograms of *PKD1* MLPA. A control sample is shown at top. The sample below has a deletion encompassing the entire *PKD1* gene, with all peaks of reduced height indicated with an asterisk. The location of each probe within the gene is shown: P1, *PKD1;* T2, *TSC2;* P2, *PKD2;* e, exon; control, control probe; i, intron; 5′flank, 5′ flanking probe. The sizes of the amplified products are 90–154 nucleotides. (B) Normalized peak height graphs for eight DNA samples. Each bar represents the normalized peak height for the probe indicated on the *x* axis, with control probes first, then *PKD1* probes, then a *TSC2* probe, then two *PDK2* probes. The heavy black lines indicate probes with reduced signal. Upper left is a control sample; next is an apparent single-exon deletion, which proved to be a single-nucleotide deletion under the probe. The remainder are all deletions extending from *TSC2* variable distances into the *PKD1* gene.

We have also demonstrated that genomic deletion mutations in *PKD1* are relatively rare. We found none among 60 ADPKD patients belonging to 27 families, 12 of whom were from a population enriched for such mutations by prior *PKD1* small mutation screening. This matches well with previous analyses of PKD patients that employed Southern blot analysis, with findings of 3 of 124 (2.4%) deletions [15]. By comparison, we recently completed a comprehensive survey for genomic deletion mutations in TSC patients and found that 6.5% of TSC patients had genomic deletion/duplication mutations in *TSC1* or *TSC2* [18]. In addition, in that study, 21 of 54 (39%) patients with such mutations had a deletion that encompassed portions of both *TSC2* and *PKD1*. The inference is that most genomic deletions occurring in *PKD1* are relatively large, extending to include the neighboring *TSC2* gene and presenting early in life with both TSC and early onset PKD [17].

It is notable that this *PKD1* MLPA probe set now enables a search for deletion mutations within *PKD1* at a small fraction of the cost, effort, and DNA quantity required for Southern blot analysis. In addition, the method that we have described for generation of MLPA probe sets that can distinguish between closely related genomic sequences, by placing variant nucleotides at the ligation site of the MLPA probe sets, can be adapted for use in the analysis of any repetitive genomic sequence as long as there are some variant nucleotides among the copies.

Although the total number of independent samples studied was small, we found no evidence for an enhanced rate of genomic deletion events near the polypurine–polypyrimidine tract present in intron 21 of *PKD1*. A more global destabilizing effect is possible, leading to deletions that extend across or are nearby this tract. However, it is difficult to provide clear evidence for this potential mechanism. In addition, we did not examine somatic mutations in *PKD1* in renal tubule cells. Such somatic mutations are thought to be a critical step in the pathogenesis of cyst development in PKD patients, since cyst-lining cells often demonstrate loss of heterozygosity at the *PKD1* locus [19,20]. The polypurine–polypyrimidine tract may contribute to somatic mutagenesis of *PKD1* and cyst development.

Materials and methods

Patient samples

Several sets of DNA samples from human subjects were available for this study, all collected after HRC-approved informed consent. Forty-eight ADPKD patients from 15 families were collected by J.B. in Cincinnati, Ohio, USA, and all met conventional diagnostic criteria for this condition. Twelve ADPKD patients were collected by Y.P. in Toronto, Ontario, Canada. These 12 patients were part of a larger set of 23 ADPKD patients who were screened for *PKD1* and *PKD2* mutations by direct sequencing (Athena Diagnostics, Inc.) and in 11 of whom mutations were identified (unpublished data). Fifteen tuberous sclerosis patients were collected by D.J.K. and collaborators and were previously shown to have genomic deletions of *TSC2* that extended into *PKD1* [18].

PKD1 MLPA design

MLPA oligonucleotide probes were designed to assess copy number at multiple sites within the *PKD1* gene. Each probe was composed of two 5′ and 3′ half-probes, each containing unique target-specific sequence, stuffer sequence, and universal primer sequences on their 5′ and 3′ ends, respectively [21]. The probe design strategy was as described previously [18] except that additional analysis was performed for probes located in the duplicated regions of *PKD1*. Candidate sequences for probes in duplicated regions were compared (BLASTN algorithm with Expect=1 and without filtering) against the reference human genome sequence. Highly homologous repeat copies were then all aligned at once using MultiAlign (http://prodes.toulouse.inra.fr/multalin/multalin.html) [22]. To maximize discrimination of *PKD1* vs these other sequences by the probe sets, probes were designed by placing variant nucleotides (between the *PKD1* sequence and the repeats) at the ligation position in 5′ or 3′ half-probes (Fig. 1B). All probes were synthesized at 100 nM scale and purified by PAGE (IDT, Skokie, IL, USA). To facilitate ligation, 3′ half-probes were synthesized with a 5′ phosphate.

MLPA reactions

MLPA was performed as described previously [18], following the general directions provided by MRC-Holland (www.mlpa.com), using a probe set to cover the entire *PKD1* gene. Briefly, 3.4 μl genomic DNA (20 ng/μl) was incubated at 98 °C for 5 min. After the sample was cooled to room temperature, 1 μl of probe mix (containing 1 and 2 fmol of probes located in unique and duplicated regions, respectively) and 1 μl of SALSA hybridization buffer were added, and the solution was denatured at 95 °C for 2 min and hybridized at 60 °C for 16 h. Hybridized probes were ligated at 54 °C for 15 min by addition of 21 μl ligation mixture. Following heat inactivation, 7.5 μl of ligation reaction was mixed with 22.5 μl of PCR buffer, heated to 60 °C, mixed with 7.5 μl PCR mixture (SALSA polymerase, dNTPs, and universal primers, one of which was labeled with fluorescein), and subjected to PCR amplification for 30 cycles. All reagents except the synthesized oligonucleotide probes were obtained from MRC-Holland.

Amplification products were diluted in water and then 1:9 in HiDi formamide (ABI) containing 1/36 volume of ROX500 size standard (ABI) (final dilution 20-fold) and then separated by size on an ABI 3100 genetic analyzer (ABI). Electropherograms were analyzed by GeneMapper version 3.5 (ABI), and peak height data were exported to an Excel table. Excel programs were generated (available upon request) to transform the peak height data to normalized values, such that control samples gave a value of 1 after normalization. Briefly, peak heights for each probe were divided by the average signal from three or more control probes (located on different chromosomes), and then that value was divided by a similar value calculated from reference samples. We used the average values from four reference samples without deletion in *TSC2*/*PKD1* processed concurrently for each analysis.

DNA sequence analysis

DNA sequence analysis was performed using conventional Big Dye Sequencing by the BWH DNA Sequencing Core Facility.

Acknowledgments

We thank Mei Lin and Dawn Ciulla for assistance with performance of MLPA capillary runs and subsequent data capture and analysis. We also thank Elizabeth Thiele, David Franz, and Sergiusz Jozwiak, for contributing TSC patient samples for this analysis, and all of the TSC and PKD patients for participating in this study. This work was supported by grants from the NIH NINDS (NS31535 to D.J.K.) and Kidney Foundation of Canada (to Y.P.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.10.003.

References

[1] O.Z. Dalgaard, Bilateral polycystic disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families, Acta Med. Scand., Suppl. 328 (1957) 1–255.

[2] G.M. Fick, P.A. Gabow, Hereditary and acquired cystic disease of the kidney, Kidney Int. 46 (1994) 951–964.

[3] D.J. Peters, L.A. Sandkuijl, Genetic heterogeneity of polycystic kidney disease in Europe, Contrib. Nephrol. 97 (1992) 128–139.

[4] P.C. Harris, et al., Polycystic kidney disease. 1. Identification and analysis of the primary defect, J. Am. Soc. Nephrol. 6 (1995) 1125–1133.

[5] The European Polycystic Kidney Disease Consortium, The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16, Cell 77 (1994) 881–894.

[6] N. Bogdanova, et al., Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes, Genomics 74 (2001) 333–341.

[7] S. Rossetti, et al., Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications, Am. J. Hum. Genet. 68 (2001) 46–63.

[8] P. Peltola, et al., Genetics and phenotypic characteristics of autosomal dominant polycystic kidney disease in Finns, J. Mol. Med. 83 (2005) 638–646.

[9] R.T. Blaszak, et al., DNA structural transitions within the PKD1 gene, Nucleic Acids Res. 27 (1999) 2610–2617.

[10] A. Bacolla, et al., Pkd1 unusual DNA conformations are recognized by nucleotide excision repair, J. Biol. Chem. 276 (2001) 18597–18604.

[11] H.P. Patel, et al., PKD1 intron 21: triplex DNA formation and effect on replication, Nucleic Acids Res. 32 (2004) 1460–1468.

[12] J.J. Bissler, Triplex DNA and human disease, Front. Biosci. 12 (2007) 4536–4546.

[13] D.F. Callen, et al., Integration of transcript and genetic maps of chromosome 16 at near-1-Mb resolution: demonstration of a “hot spot” for recombination at 16p12, Genomics 29 (1995) 503–511.

[14] P.C. Harris, et al., A large duplicated area in the polycystic kidney disease 1 (PKD1) region of chromosome 16 is prone to rearrangement, Genomics 23 (1994) 321–330.

[15] Y. Ariyurek, et al., Large deletions in the polycystic kidney disease 1 (PKD1) gene, Hum. Mutat. 23 (2004) 99.

[16] P.T. Brook-Carter, et al., Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome, Nat. Genet. 8 (1994) 328–332.

[17] J.R. Sampson, et al., Renal cystic disease in tuberous sclerosis: role of the polycystic kidney disease 1 gene, Am. J. Hum. Genet. 61 (1997) 843–851.

[18] P. Kozlowski, et al., Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype–phenotype correlations, Hum. Genet. 121 (2007) 389–400.

[19] F. Qian, et al., The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I, Cell 87 (1996) 979–987.

[20] J.L. Brasier, E.P. Henske, Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss-of-function model for cyst pathogenesis, J. Clin. Invest. 99 (1997) 194–199.

[21] J.P. Schouten, et al., Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification, Nucleic Acids Res. 30 (2002) e57.

[22] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic Acids Res. 16 (1988) 10881–10890.